



New Measures to Access the Quality of Race/Ethnicity Reporting in State Databases

David Zingmond, MD, PhD
AHRQ monthly grantees call
January 10, 2012



Aim

- Develop validated audit measures for race/ethnicity reporting that can be used for any state's statewide databases



Background: Data Auditing

- Current data auditing rules are blunt instruments for determining the accuracy (or adequacy) of race/ethnicity reporting in statewide hospital databases
 - Provide minimum criteria for flagging hospital data
 - Flags for hospitals exceeding rates for missing or unknown race/ethnicity
 - Flags for extreme variation in reporting (100% or 0% for categories)



Unknown race/ethnicity

- Combined race/ethnicity measure
- 3.4% unknown race/ethnicity (mean across 349 hospitals reporting discharges in 2009)
 - 17 hospitals > 10% unknown
 - 1 hospital > 20% unknown



Candidate Audit Measures

- Reflect self-reported race/ethnicity
- Data are readily available for use by those performing the data audits
- Patient-level record comparisons are best
 - not available for every patient or for every state (e.g. where patient-level data linkages are not routinely performed)



Available Data for Comparison

- Patient-level data
 - Birth certificates (mother self-report; not all hospitals have births)
 - Cancer registry (abstracted with data supplemented by name-based algorithm and death certificates; small subset with self-report)
 - Death certificates (institutional reporting)
- U.S. Census data (self-report)



Measures for Agreement – Patient Level Data

- Patient-level data

$$\% \text{ agreement} = \# \text{ agree with GS} / \# \text{ Total} * 100$$

- This type of measure can be used for single category agreement or overall agreement across all categories
- If GS (gold standard) is truly self-report, then this measure can also be used for validation purposes



Measures for Agreement – Summary Level Data

- Hospital Summary Level Data

Estimated disagreement =

$\text{Abs}(\text{Reported} - \text{Predicted})/2 + \% \text{ Unknown}$

Reported = Distribution of race categories

Predicted = Population mean predicted
distribution using zip-code level distribution
for each patient

* agreement = 100 – disagreement



PDD vs CCR

	PDD vs. SR	PDD vs. CCR
Race	0.91	0.92
Ethnicity	0.91	0.95
Alt Ethnicity (unk = NH)	0.95	0.95
White-White	0.94	0.93
Non-White-Non-White Race	0.82	0.91
Hispanic-Hispanic	0.66	0.66
	PDD vs. SR (combined race/ethnicity)	PDD vs. CCR (combined race/ethnicity)
Combined Race/Ethnicity	0.90	0.92
NH White-NH White	0.95	0.96
Non-White-Non-White	0.79	0.83
Hispanic-Hispanic	0.66	0.66

N = 16,653

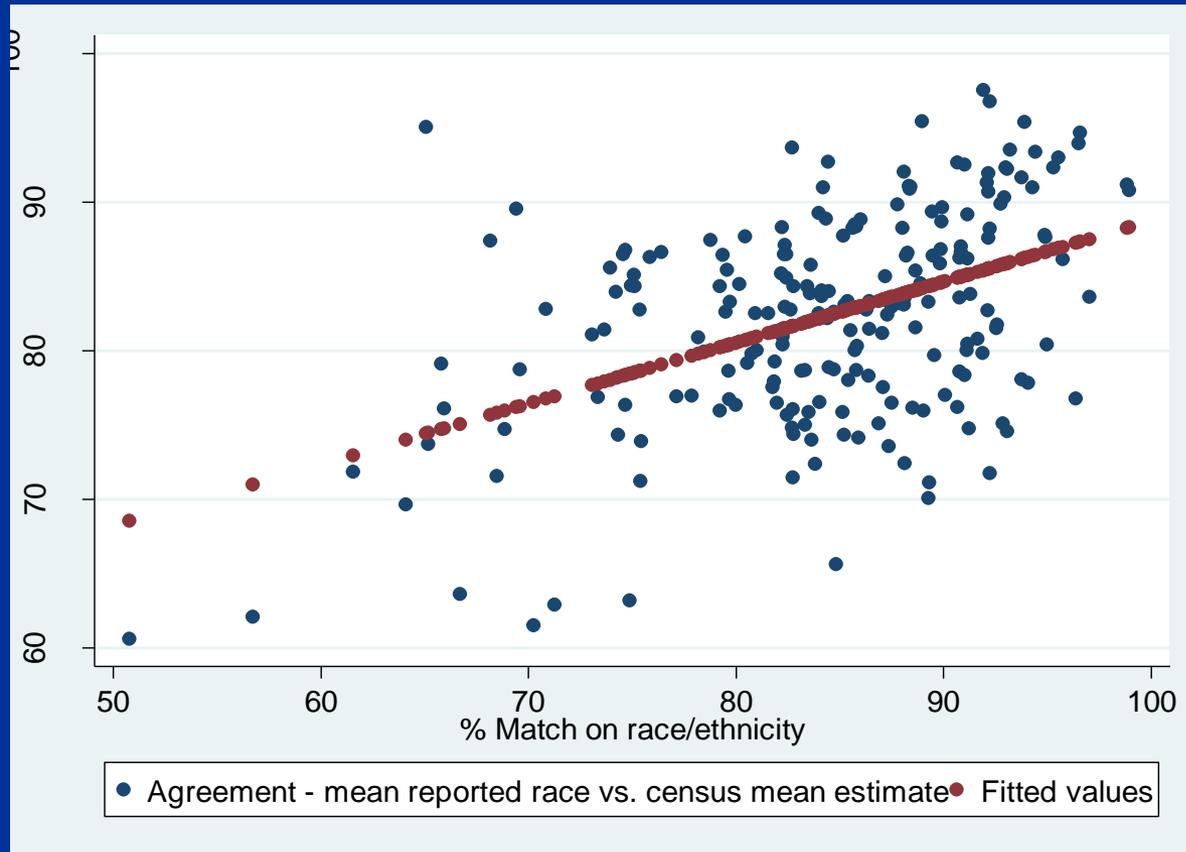


PDD versus Birth

Race	70.0
Ethnicity	85.8
White	71.6
NW	65.4
Hisp	89.1
NH	82.1
Race/Eth	85.8
NHW	91.1
minorities	83.9

N = 513,456

Agreement (maternal) vs Mean Population Agreement



$$\rho = 0.5$$

N = 513,456 at 254 Hospitals where births occurred;
Agreement between PDD & birth versus PDD & Census



Other observations

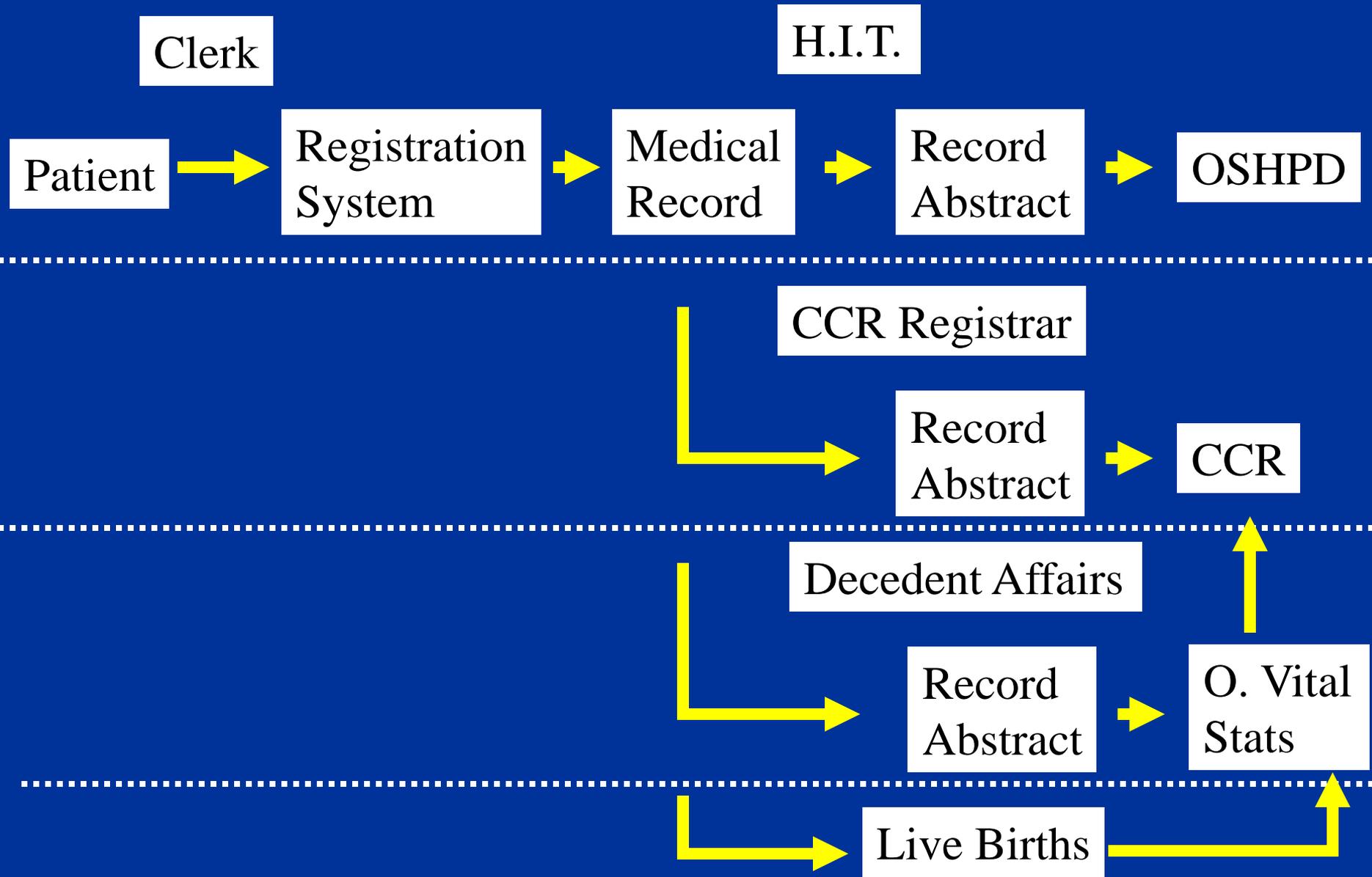
- Too much scatter for good matching for prediction
- Populations are not exact matches (mothers versus all adult patients)



Further work on Data Audits

- Revised metrics
- Match populations for derivation of metrics
- Compare proposed metrics to current insensitive, context-free metrics
- Apply metrics to hospitals across time to see if there have been changes in performance during the observation period

Data Collection and Reporting



Existing Gold Standard for Self-Report

